Cross-corpus Feature Learning between Spontaneous Monologue and Dialogue for Automatic Classification of Alzheimer's Dementia Speech

Cross-corpus Feature Learning between Spontaneous Monologue and Dialogue for Automatic Classification of Alzheimer's Dementia Speech

Sofia de la Fuente Garcia[1], Fasih Haider[1], Saturnino Luz[1]

***Abstract*— Speech analysis could help develop clinical tools for automatic detection of Alzheimer's disease and monitoring of its progression. However, datasets containing both clinical information and spontaneous speech suitable for statistical learning are relatively scarce. In addition, speech data are often collected under different conditions, such as monologue and dialogue recording protocols. Therefore, there is a need for methods to allow the combination of these scarce resources. In this paper, we propose two feature extraction and representation models, based on neural networks and trained on monologue and dialogue data recorded in clinical settings. These models are evaluated not only for AD recognition, but also with respect to their potential to generalise across both datasets. They provide good results when trained and tested on the same data set (72.56% UAR for monologue data and 85.21% for dialogue). A decrease in UAR is observed in transfer training, where feature extraction models trained on dialogues provide better average UAR on monologues (63.72%) than the other way around (58.94%). When the choice of classifiers is independent of feature extraction, transfer from monologue models to dialogues result in a maximum UAR of 81.04% and transfer from dialogue features to monologue achieve a maximum UAR of 70.73%, evidencing the generalisability of the feature model.**

Clinical relevance **We present a method for automatic screening of cognitive health in dementia risk settings. The method is based on spoken language, an ubiquitous source of data, therefore being cost-efficient, non-invasive and with little infrastructure required.**

I. INTRODUCTION

Dementia is a category of neurodegenerative diseases that entails a long-term decrease of cognitive functioning. Gradually, the severity of the symptoms (i.e. memory loss, thought difficulties, language impairment, motor problems, emotional distress) increases at the expense of the patient's autonomy, as well as their well-being and their caregivers' [1]. Those cognitive symptoms may be a consequence of the neuropathology that starts silently up to 20 years before they become observable, with currently no satisfactory treatment.

In 2015, the WHO [2] estimated approximately 47.5 million cases of dementia worldwide, confirming the severity of the situation and anticipating a massive societal impact. Therefore, there is a need for cost-effective and scalable methods ready to recognise dementia from pre-clinical and mild stages to Alzheimer's Dementia (AD).

The symptomatic heterogeneity of AD demands diagnostic methods that are able to capture subtler and broader aspects than conventional screening tools (e.g. Mini-Mental State Examination [3]), which often fail to discriminate pre-clinical symptoms. In addition, signal processing technology is creating opportunities for personal health monitoring and development of diagnostic support based on automated processing of behavioural signals [4]. Whilst the most prominent of these signals for AD is frequently considered to be memory loss, language alterations also appear early in the disease [5]. This, together with the fact that speech and language are rich and ubiquitous sources of cognitive behavioural data, provide computational technologies with a broad potential for contribution as diagnostic-support tools [6].

Although spontaneous and longitudinal speech data would be ideal for cognitive screening and disease monitoring, conventional cognitive assessments evaluate speech and language under controlled laboratory conditions. Despite there being an increasing tendency to collect "real life" speech data, so far there are few available datasets, two of which are used in this study. The first one, *The Pitt Corpus*, consists of spontaneous narrative speech (monologues) from participants with various degrees of AD. The second one, the *Carolina Conversations Collection* (*CCC*), is amongst the few spontaneous dialogue datasets available in the context of AD research.

In this paper, we demonstrate a machine learning approach for AD recognition with acoustic information extracted from spontaneous speech. We propose two Feature Extraction Models (FEMs), based on neural networks and trained on monologue (*The Pitt Corpus*(and dialogue (*CCC*) data. These FEMs are not only evaluated for AD recognition, but also with respect to their potential to generalise across both datasets. As dialogues and monologues differ at various linguistic levels [7], the ability to "transfer" FEMs across these two types of speech data will be advantageous for future research and clinical applicability. Given the scarcity of speech data for AD classification research, the enhanced and pre-processed versions of these two datasets are an additional contribution to the field. These will be available upon request and could serve as benchmark datasets for the research community.

II. BACKGROUND

A variety of computational methods have been applied to attempt detection of AD or MCI on both of the aforementioned datasets. One of the most comprehensive models for was trained on *the Pitt corpus* (monologue speech) and achieved 81.92% accuracy for machine learning classification of individuals with and without AD [8]. Unlike our work,

[1] S. de la Fuente Garcia, F. Haider and S. Luz are with the Usher Institute, Edinburgh Medical School, The University of Edinburgh, UK, {sofia.delafuente, fasih.haider, s.luz}@ed.ac.uk

D. la Fuente Garcia, S., Haider, F., and Luz, S. Cross-corpus feature learning between spontaneous monologue and dialogue for automatic classification of Alzheimer's dementia speech. In *Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)* (2020), pp. 5851–5855.

most of these studies rely largely, or exclusively, on high-level linguistic features derived from the manual transcriptions available with the speech data (e.g. [9], [10], [11], [12]).

Dementia research is incrementally investigating spontaneous conversations as a source of clinical information to support diagnostic protocols. The state-of-the-art model trained on dialogue data achieved 90.9% accuracy in binary classifications between healthy controls and neurodegenerative memory disorder [13]. However, the performance of this model drops to 68% when trying to detect MCI and it was not tested on a group of AD patients [14].

Conversational research in the context of AD is far more limited than monologue research. A study on the *CCC* corpus reported 80% precision and recall with a Naive Bayes classifier between AD from non-AD speech. They used transcription-derived linguistic metrics and pragmatic dialogue features [10]. More recently, a study set the state-of-the-art for dialogue AD research by obtaining 85% accuracy on *CCC*, with an additive logistic regression model. They focused on paralinguistic features of dialogue by extracting graph-based features encoding turn-taking patterns and speech rate from dialogues involving an AD speaker and from non-AD dialogues [15].

Paralinguistic approaches to AD classification have been less researched so far, but strong arguments support their investigation. On the one hand, there are methodological reasons, such as avoiding the constrains inherent to transcription procedures. On the other hand, acoustic analysis may contribute to our understanding of the disease by pointing out speech subtleties that could have a localised neural substrate. Although the comprehensive model by [8] included acoustic features, only two recent studies have relied exclusively on audio recordings from *the Pitt Corpus*, and used them to train a classifier for AD detection. The first one, obtained a 68% accuracy by training a Bayesian classifier solely with low-level acoustic features (vocalisation events, speech rate and number of utterances over a discourse events) [16]. Subsequently, they increased this to 78.7% by using standardised feature sets (emobase, eGeMAPS, ComParE) and several different machine learning classifiers [17].

To the best of our knowledge, these are the only studies on *the Pitt Corpus* to exclusively employ speech data – without relying on transcripts – in order to classify AD patients and elderly controls. Furthermore, [17] is the first attempt to use standardised paralinguistic feature sets in this context. As for *CCC*, only one study has worked directly with the recordings, extracting MFCC and linguistic features. They achieved 79.5% accuracy when classifying utterances with and without "trouble indicating behaviours" for AD [18].

In a nutshell, the application of speech technology to dementia research is a heterogeneous field in which comparisons are difficult to establish. The work hereby presented aims to assess whether a FEM trained on a balanced version of the *Pitt Corpus*, generalises well to an enhanced version of the *CCC*, and vice versa. In other words, whether our proposed model is able to generalise across these monologue and dialogue speech datasets.

III. METHODOLOGY

A. The Pitt Corpus

The Pitt Corpus[1] was gathered longitudinally at the University of Pittsburgh and distributed through DementiaBank [19]. Participants, over 44 years old, undertook extensive neuropsychological and physical assessments and were categorised into three diagnostic groups: AD, healthy controls (HC) and unkwnown [19]. HC and AD groups were recorded while performing The Cookie Theft Picture description task [20], which generates spontaneous narrative speech linked to neuropsychological data. In order to to minimise risk of bias in classification results, we created a derived dataset matched for age and gender. The resulting dataset was pre-processed for enhancement, and segmented for voice activity. Our final experimental dataset contains 2033 speech segments from 82 non-AD subjects and 2043 speech segments from 82 AD subjects (46 females).

B. Carolina Conversations Collection

CCC[2], hosted and distributed by the Medical University of South Carolina, is a digital collection conversations (including both voice recordings and transcriptions) about health with patients, over 65 years old, suffering various chronic health conditions [21]. Participants are labeled by diagnosis, and for the purposes of our experiments we created an AD group, which included participants diagnosed with AD, and a non-AD group, which included participants with other chronic conditions (e.g. diabetes, heart disease). Different to *The Pitt Corpus*, *CCC* does not provide clinical information other than the participants' diagnoses.

Due to the size of the dataset (30 participants in the AD group and 16 in the non-AD group), we did not match by age and gender, since it would have significantly reduced the number of instances. The selected recordings were also enhanced and segmented, resulting in 9,354 dialogue instances, from 80 conversations that belong to 30 participants (23 females) in the AD group. The non-AD group contains 7,052 dialogue instances, from 139 conversations that belong to 16 participants (14 females).

C. Pre-processing

Since both datasets present undermined quality for acoustic analysis, we implemented three pre-processing steps. First, stationary random noise was estimated for each audio file and removed from the speech spectrum (i.e. spectral subtraction) [22]. Second, in order to control for variations caused by recording conditions, audio volume was normalised across files by applying a constant amount of gain to the entire recordings for the amplitude to reach a certain level (norm), whilst the signal to noise ratio and relative dynamics remain unchanged. Third, a voice activity detection system based on signal energy threshold [23] and the timestamps provided were used for speech segmentation. *The Pitt Corpus* was further segmented to remove long silences in

[1]https://dementia.talkbank.org/

[2]https://carolinaconversations.musc.edu

D. la Fuente Garcia, S., Haider, F., and Luz, S. Cross-corpus feature learning between spontaneous monologue and dialogue for automatic classification of Alzheimer's dementia speech. In *Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)* (2020), pp. 5851–5855.

order to even out the speech segments for a more comparable feature extraction. This was not necessary for CCC. The segmentation process increased the sample size in terms of number of instances available for analysis, even though the number of participants is not that large. We used the enhanced recordings to train the machine learning model described below. Enhanced datasets and the code for their pre-processing are available upon request.

D. Feature Extraction Model (FEM)

The following procedure was applied to both datasets. First, acoustic feature extraction was performed on the speech segments using the openSMILE v2.1 toolkit [24]. We extracted the *eGeMAPS* [25] feature set. This feature set contains the F0 semitone, loudness, spectral flux, MFCC, jitter, shimmer, F1, F2, F3, alpha ratio, Hammarberg index and slope V0 features, as well as their most common statistical functionals, for a total of 88 features per speech segment.

Subsequently, we applied Active Data Representation (ADR), a method we have recently proposed [26], [17], [27]. ADR considers the acoustic features extracted from all the speech segments of an audio recording and represents them with a single fixed-dimension feature vector for the classification task. The term '*Active* Data Representation' is used because we did not use any event detector (such as emotion recognition) at a speech segment level for generating a feature vector for AD classification [28]. ADR models acoustic information accounting for different recordings produced by the same subject, granting subject independence. This is the first time cross-corpus transfer is attempted with ADR. Generating the ADR involves the following steps:

1) *Segmentation and feature extraction*: each audio recording Ai ($i = 1 \dots N$, where N represents the total number of audio recordings or subjects) is divided into n speech segments $S_{k,Ai}$, where k varies from 1 to n. Hence $S_{k,Ai}$ is the k^{th} segment of the i^{th} audio recording, and acoustic features are extracted over such speech segments, rather than over the full audio recording, at this processing stage.
2) *Clustering of segments*: self-organising maps (SOM) [29] are employed for clustering segments $S_{k,Ai}$ into m clusters $(C_1, C_2, ..., C_m)$ using audio features. SOM is an attractive clustering method in this context as it addresses both topology and distribution, and requires no assumptions regarding the input vectors. Furthermore, it has been previously used for speech segment clustering based on voice styles with good results [30], [31]. Here m represents the number of SOM clusters that correspond to the FEM. The number of clusters is determined through grid search over $m \in \{5, 10, \dots, 100\}$.
3) *Generation*: Active Data Representation (ADR_{Ai}) vectors are generated by first computing the number of segments in each cluster for each audio recording (Ai), that is, creating a histogram representation of the number of speech segments ($nADR_{Ai}$) present in each of the m clusters for each audio recording. Then, to model temporal dynamics the mean and standard deviation of the rate of change with respect to the sizes of the clusters of speech segments for each audio recording ($vADR_{Ai}$) is calculated. Finally, a histogram representation of segment duration ($dADR_{Ai}$) is built for each cluster Ai.
4) *Normalisation*: as the number and duration of segments is typically different for each audio recording or subject due to inter-subject variability, we normalise the feature vector by dividing it by the L1 norm of $nADR_{Ai}$ and $dADR_{Ai}$, respectively.
5) *Feature Fusion*: the $ADR_{Ai_{norm}}$ feature set encompasses the features of $nADR_{Ai_{norm}}$, $dADR_{Ai_{norm}}$ and $vADR_{Ai}$. Therefore a feature vector using feature (early) fusion with dimensionality of $2 \times (m + 1)$ is generated to represent each subject.

E. Classification Methods

The classification experiments were performed using five different methods, namely decision trees (DT, with leaf size of 20), nearest neighbour (KNN with K=1), linear discriminant analysis (LDA) and support vector machines (SVM, with a linear kernel with box constraint of 0.1, and sequential minimal optimization solver). The classification methods are implemented in MATLAB using the statistics and machine learning toolbox [32]. More sophisticated classification models (e.g. recurrent structures) were discarded in order to avoid the risk of over-fitting due to the relatively small number of subjects. A leave-one-subject-out (LOSO) cross-validation setting was adopted, where the training data do not contain any information of validation subjects. To assess the classification results, we used unweighted average recall (UAR) instead of overall accuracy since the CCC data set is imbalanced. The unweighted average recall is the arithmetic average of recall of all classes.

F. Experimentation

We conducted two experiments. In the first (hereafter $Pitt_{ADR}$) we employ the LOSO procedure to train FEM using Pitt data, generate $ADR_{Ai_{norm}}$ and then train and test on Pitt. Then, we map CCC data through the FEM based on Pitt data and generate $ADR_{Ai_{norm}}$ for training and testing (LOSO) on CCC subjects. The parameter m for$ADR_{Ai_{norm}}$ is optimised on the results of Pitt subjects and validated on CCC subjects.

In the second experiment (CCC_{ADR}), we train FEM on CCC data, generate $ADR_{Ai_{norm}}$ and classify CCC subjects (again, employing LOSO cross-validation). Then, Pitt data are mapped onto the CCC FEM to generate $ADR_{Ai_{norm}}$, which is used for classifying Pitt subjects. The parameter m for $ADR_{Ai_{norm}}$ is optimised on the results of CCC subjects and validated on Pitt subjects.

IV. RESULTS AND DISCUSSION

The results of the experimentation are shown in Table I. It is noted that the $Pitt_{ADR}$ provides the best UAR (72.56%)

D. la Fuente Garcia, S., Haider, F., and Luz, S. Cross-corpus feature learning between spontaneous monologue and dialogue for automatic classification of Alzheimer's dementia speech. In *Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)* (2020), pp. 5851–5855. © 2020 IEEE.

using the DT classifier with a value of 100 for the m parameter for FEM. However, testing this FEM ($m = 100$) with CCC on the same classifier results in a decreased UAR (53.12%). This could be due to the imbalanced nature of *CCC*, which is not handled well by the classification algorithm (DT). On the other hand, where $Pitt_{ADR}$ provides the least UAR (51.83%), using the NB (a classifier that is more robust to class imbalance) and $m = 75$ for FEM, the performance on *CCC* rises to 81%. The confusion matrix with precision, recall, overall accuracy and Kappa [33] is shown in Figure 1. The averaged UAR (i.e. 58.94% for *Pitt Corpus* and 61.63% for *CCC*) indicated that a FEM trained using *Pitt Corpus* can extract discriminating features from the *CCC* dataset.

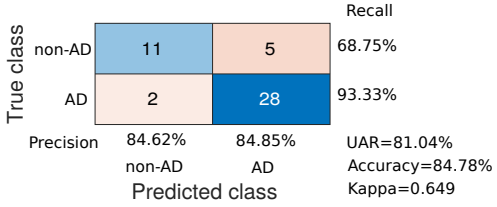

Fig. 1. Confusion matrix of the best result on *CCC* data obtained using transfer learning from *Pitt* to *CCC* data.)

As regards the second experiment, CCC_{ADR} provides the best UAR (85.21%) using the DT classifier, for a FEM where $m = 25$. Mapping Pitt features to this FEM results in a UAR of 65.24%. It is also noted that the CCC_{ADR} provides a UAR (68.33%) using the RF classifier with a value of 25 for m parameter for FEM. However, when we test this FEM ($m = 25$) with RF on the *Pitt Corpus* we see an increase in UAR to 70.73%. The confusion matrix with precision, recall, overall accuracy and Kappa [33] is shown in Figure 2. Together with the results of the first experiments, this indicates that, even though feature learning transfers well between the two data sets, the choice of classifier must be problem specific. Overall, the mean UAR (i.e. 63.72% for Pitt and 76.87% for *CCC*) indicated that a FEM trained using *CCC* generalizes reasonably well to Pitt data. Based on the above findings it could be argued that dialogue data (*CCC*) offers a better platform for FEM training and transfer learning than monologue data (*Pitt Corpus*) for our particular research purposes. While there is a decrease in performance relative to FEM trained and tested on the same data (72.56% UAR for *Pitt* and 85.21% for *CCC*), the proposed transfer method can be effective if decoupled from classifier choice.

TABLE I

UAR (%) FOR BOTH EXPERIMENTS.

	$Pitt_{ADR}$			CCC_{ADR}		
Classifier	**m**	**Pitt**	CCC	**m**	Pitt	**CCC**
DT	100	*72.56*	**53.12**	25	**65.24**	*85.21*
LDA	35	55.49	63.12	20	68.29	82.08
KNN	45	56.10	63.33	10	62.80	82.50
SVM	40	58.54	53.12	15	68.29	60.83
NB	75	*51.83*	**81.04**	35	46.95	82.29
RF	100	59.15	56.04	25	**70.73**	*68.33*
Mean	–	58.94	61.63	–	***63.72***	***76.87***

The cross-corpus experiments reported here offer insights into the potential for transfer learning (in the specific sense described above) between dialogue and monologue based models. Furthermore, our results suggest better transference from dialogue to monologue than the other way around. Hence, they align with the the psycholinguistics hypothesis that dialogues encompass a wider range of prosodic aspects than monologues [7].

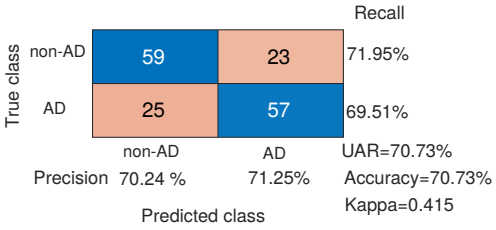

Fig. 2. Confusion matrix of the best result on Pitt data obtained using transfer learning from CCC to Pitt data.)

V. CONCLUSIONS

The conclusions to be drawn from this paper are twofold. First, we demonstrate that spontaneous speech is a valuable source of information for AD recognition, both in monologue (narrative) and dialogue (conversational) format. Also, we show how our ADR-based method for cross-corpus learning performs better with a FEM trained on dialogues to be used for monologue classification than the other way around. This occurs even though the FEM trained on monologues obtained a better accuracy when tested on these monologues, than the accuracy obtained by the dialogue FEM on the dialogues.

As a limitation of the study, it is worth noting that there are further differences between datasets, aside from recording protocol (i.e. monologue vs. dialogue), such as recording conditions or devices. This needs to be accounted for in order to be able to guarantee that differences in classifier performance are caused by the type of speech.

In future research we aim to extend the work presented in this paper by incorporating Mini-mental State Examination scores [34] available in the *Pitt Corpus*, and evaluating the potential of our model to predict them. In addition, we are currently collecting naturalistic dialogue data [35] where participants are healthy adults at risk of AD, along with comprehensive genetic, cognitive and family history, imaging and biomarker data. The results presented here support our hypothesis that dialogue might be a better suited source of acoustic data for early detection of AD. Hence, we look forward to implementing this procedure on other spontaneous dialogue data and extending the research to include neuropsychological assessment.

ACKNOWLEDGMENTS

This research has received funding from the European Union's Horizon 2020 research programme, under grant agreement No 769661, towards the SAAM project. S. de la Fuente Garcia is supported by the Medical Research Council (MRC), UK. We acknowledge B. MacWhinney (DementiaBank), C. Pope and B.H. Davis (CCC) for hosting and sharing the databases used in this research.

D. LA FUENTE GARCIA, S., HAIDER, F., AND LUZ, S. Cross-corpus feature learning between spontaneous monologue and dialogue for automatic classification of Alzheimer's dementia speech. In *Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)* (2020), pp. 5851–5855. © 2020 IEEE.

REFERENCES

[1] American Psychiatric Association, "Delirium, dementia, and amnestic and other cognitive disorders," in *Diagnostic and Statistical Manual of Mental Disorders, Text Revision (DSM-IV-TR)*, 4th ed., American Psychiatric Association, Ed., Arlington, VA, 2000, ch. 2.

[2] World Health Organization, "First WHO ministerial conference on global action against dementia: meeting report," *WHO Library Cataloguing-in-Publication DataLibrary Cataloguing-in-Publication Data*, pp. 1–76, 2015.

[3] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "Mini-mental state. A practical method for grading the cognitive state of patients for the clinician." *Journal of psychiatric research*, vol. 12, no. 3, pp. 189–98, 1975.

[4] P. N. Dawadi, D. J. Cook, and M. Schmitter-Edgecombe, "Automated cognitive health assessment using smart home monitoring of complex tasks," *IEEE transactions on systems, man, and cybernetics: systems*, vol. 43, no. 6, pp. 1302–1313, 2013.

[5] G. W. Ross, J. L. Cummings, and D. F. Benson, "Speech and language alterations in dementia syndromes: Characteristics and treatment," *Aphasiology*, vol. 4, no. 4, pp. 339–352, 1990.

[6] A. J. Braaten, T. D. Parsons, R. Mccue, A. Sellers, and W. J. Burns, "Neurocognitive Differential Diagnosis Of Dementing Diseases: Alzheimer's Dementia, Vascular Dementia, Frontotemporal Dementia, And Major Depressive Disorder," *International Journal of Neuroscience*, vol. 116, no. 11, pp. 1271–1293, 2006.

[7] M. J. Pickering and S. Garrod, "Toward a mechanistic psychology of dialogue," *Behavioral and Brain Sciences*, vol. 27, no. 02, 2004.

[8] K. C. Fraser, J. A. Meltzer, and F. Rudzicz, "Linguistic features identify Alzheimer's disease in narrative speech," *Journal of Alzheimer's Disease*, vol. 49, no. 2, pp. 407–422, 2016.

[9] C. I. Guinn and A. Habash, "Language analysis of speakers with dementia of the Alzheimers type," in *2012 AAAI Fall Symposium Series*, 2012.

[10] C. Guinn, B. Singer, and A. Habash, "A comparison of syntax, semantics, and pragmatics in spoken language among residents with Alzheimer's disease in managed-care facilities," in *2014 IEEE Symposium on Computational Intelligence in Healthcare and e-health (CICARE)*. IEEE, 2014, pp. 98–103.

[11] R. Ben Ammar and Y. Ben Ayed, "Speech Processing for Early Alzheimer Disease Diagnosis: Machine Learning Based Approach," in *2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA)*, 2018, pp. 1–8. [Online]. Available: https://ieeexplore.ieee.org/document/8612831/

[12] S. O. Orimaye, J. S. Wong, K. J. Golden, C. P. Wong, and I. N. Soyiri, "Predicting probable Alzheimers disease using linguistic deficits and biomarkers," *BMC bioinformatics*, vol. 18, no. 1, p. 34, 2017.

[13] B. Mirheidari, D. Blackburn, K. Harkness, T. Walker, A. Venneri, M. Reuber, and H. Christensen, "An avatar-based system for identifying individuals likely to develop dementia," in *Interspeech 2017*. ISCA, 2017, pp. 3147–3151.

[14] B. Mirheidari, D. Blackburn, R. OMalley, T. Walker, A. Venneri, M. Reuber, and H. Christensen, "Computational cognitive assessment: Investigating the use of an intelligent virtual agent for the detection of early signs of dementia," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 2732–2736.

[15] S. Luz, S. de la Fuente, and P. Albert, "A method for analysis of patient speech in dialogue for dementia detection," in *Resources for processing of linguistic, paralinguistic and extra-linguistic data from people with various forms of cognitive impairment*, D. Kokkinakis, Ed. ELRA, May 2018, pp. 35–42.

[16] S. Luz, "Longitudinal monitoring and detection of Alzheimer's type dementia from spontaneous speech data," in *Computer-Based Medical Systems (CBMS), 2017 IEEE 30th International Symposium on*. IEEE, 2017, pp. 45–46.

[17] F. Haider, S. de la Fuente, and S. Luz, "An assessment of paralinguistic acoustic features for detection of alzheimer's dementia in spontaneous speech," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 272–281, 2020.

[18] F. Rudzicz, L. Chan Currie, A. Danks, T. Mehta, and S. Zhao, "Automatically identifying trouble-indicating speech behaviors in Alzheimer's disease," in *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility*. ACM, 2014, pp. 241–242.

[19] J. T. Becker, F. Boiler, O. L. Lopez, J. Saxton, and K. L. McGonigle, "The Natural History of Alzheimer's Disease," *Archives of Neurology*, vol. 51, no. 6, p. 585, 1994.

[20] H. Goodglass and E. Kaplan, "The assessment of aphasia and related disorders," Philadelphia, 1983.

[21] C. Pope and B. H. Davis, "Finding a balance: The carolinas conversation collection," *Corpus Linguistics and Linguistic Theory*, vol. 7, no. 1, pp. 143–161, 2011.

[22] N. Upadhyay and A. Karmakar, "Speech enhancement using spectral subtraction-type algorithms: A comparison and simulation study," *Procedia Computer Science*, vol. 54, pp. 574–584, 2015.

[23] T. Giannakopoulos, "pyaudioanalysis: An open-source python library for audio signal analysis," *PloS one*, vol. 10, no. 12, 2015.

[24] F. Eyben, M. Wöllmer, and B. Schuller, "openSMILE: the Munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 2010, pp. 1459–1462.

[25] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. André, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan *et al.*, "The Geneva minimalistic acoustic parameter set GeMAPS for voice research and affective computing," *IEEE Transactions on Affective Computing*, vol. 7, no. 2, pp. 190–202, 2016.

[26] F. Haider, M. Koutsombogera, O. Conlan, C. Vogel, N. Campbell, and S. Luz, "An active data representation of videos for automatic scoring of oral presentation delivery skills and feedback generation," *Frontiers in Computer Science*, vol. 2, p. 1, 2020.

[27] F. Haider, P. Albert, and S. Luz, "Automatic recognition of low-back chronic pain level and protective movement behaviour using physical and muscle activity information," in *15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020) Buenos Aires, Argentina*. IEEE, 2020.

[28] F. Haider, S. De La Fuente, P. Albert, and S. Luz, "Affective speech for Alzheimer's dementia recognition." in *LREC: Resources and Process-Ing of linguistic, para-linguistic and extra-linguistic Data from people with various forms of cognitive/psychiatric/developmental impairments (RaPID)*, 2020.

[29] T. Kohonen, "The self-organizing map," *Neurocomputing*, vol. 21, no. 1-3, pp. 1–6, 1998.

[30] E. Vanmassenhove, J. P. Cabral, and F. Haider, "Prediction of emotions from text using sentiment analysis for expressive speech synthesis." in *SSW*, 2016, pp. 21–26.

[31] E. Székely, J. P. Cabral, P. Cahill, and J. Carson-Berndsen, "Clustering expressive speech styles in audiobooks using glottal source parameters," in *Twelfth Annual Conference of the International Speech Communication Association*, 2011.

[32] MATLAB, *version 9.6 (R2019a)*. Natick, Massachusetts: The MathWorks Inc., 2019.

[33] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *biometrics*, pp. 159–174, 1977.

[34] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "Mini-mental state: a practical method for grading the cognitive state of patients for the clinician," *Journal of psychiatric research*, vol. 12, no. 3, pp. 189–198, 1975.

[35] S. de la Fuente Garcia, C. Ritchie, and S. Luz, "Protocol for a conversation-based analysis study: Prevent-ED investigates dialogue features that may help predict dementia onset in later life," *BMJ Open*, vol. 9, no. 3, 2019.

D. la Fuente Garcia, S., Haider, F., and Luz, S. Cross-corpus feature learning between spontaneous monologue and dialogue for automatic classification of Alzheimer's dementia speech. In *Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)* (2020), pp. 5851–5855.